



Coursera Capstone

IBM Applied Data Science Capstone

Overview for Opening or Investing in a Restaurant in Detroit, Michigan

By: Christopher Liedel, May 22, 2020

Introduction

Detroit, the historic “Motor City,” is a major metropolis with a vibrant downtown district and interesting neighborhood communities. It hosts some of the world’s best hospitals, universities, museums, restaurants, sports arenas, music and theater events in the United States.

Affordable real estate and a talented workforce in combination with the business incentives available in designated opportunity zones, makes Detroit a great place for entrepreneurs and investors alike to call home.

For the start-up restaurateur or the seasoned investor, a first understanding of the big picture is a must. What are the popular dining establishments in each neighborhood? What are the most promising neighborhoods for development or investment regarding unmet demand and competition? This project will address these questions and begin to point a way for stakeholder decisions toward maximizing success.

Data

To answer the aforementioned business questions, analysis of Detroit neighborhoods and venues will be conducted through a process of web scraping, data wrangling and the creation of appropriate data frames; clustering and visualization techniques using the following sources:

- Google Maps of Detroit Neighborhoods for list of neighborhood names and areas.
- List of Detroit Zip Codes and geocoordinates: <https://www.public.opendatasoft.com>

- Foursquare API Developer Access: <https://foursquare.com> . For exploring venue data related to dining establishments, their popularity and ranking.
- Detroit Quick Facts for demographic data from the US Census Bureau at www.uscensus.gov

Methodology

The methodology includes:

- Data retrieval, exploration and wrangling
- Segmenting neighborhoods and venues using the K-Means clustering algorithm
- Visualization of the neighborhood and venue segments

Data Retrieval, Exploration and Wrangling

1. Demographic data of Detroit was retrieved from the US Census Bureau Quick Facts at www.census.gov. It contains population, age, sex, income, housing and business data from 2010 to 2019.

The dataset was processed into a Pandas DataFrame where columns were renamed, and unnecessary columns and rows were dropped. The DataFrame was then saved to a new csv file.

		Fact	Detroit
0	Population estimates, July 1, 2019, (V2019)		670,031
1	Population estimates base, April 1, 2010, (V2010)		713,898
2	Population, percent change - April 1, 2010 (estimated)		-6.1%
3	Population, Census, April 1, 2010		713,777
4	Persons under 5 years, percent		7.3%

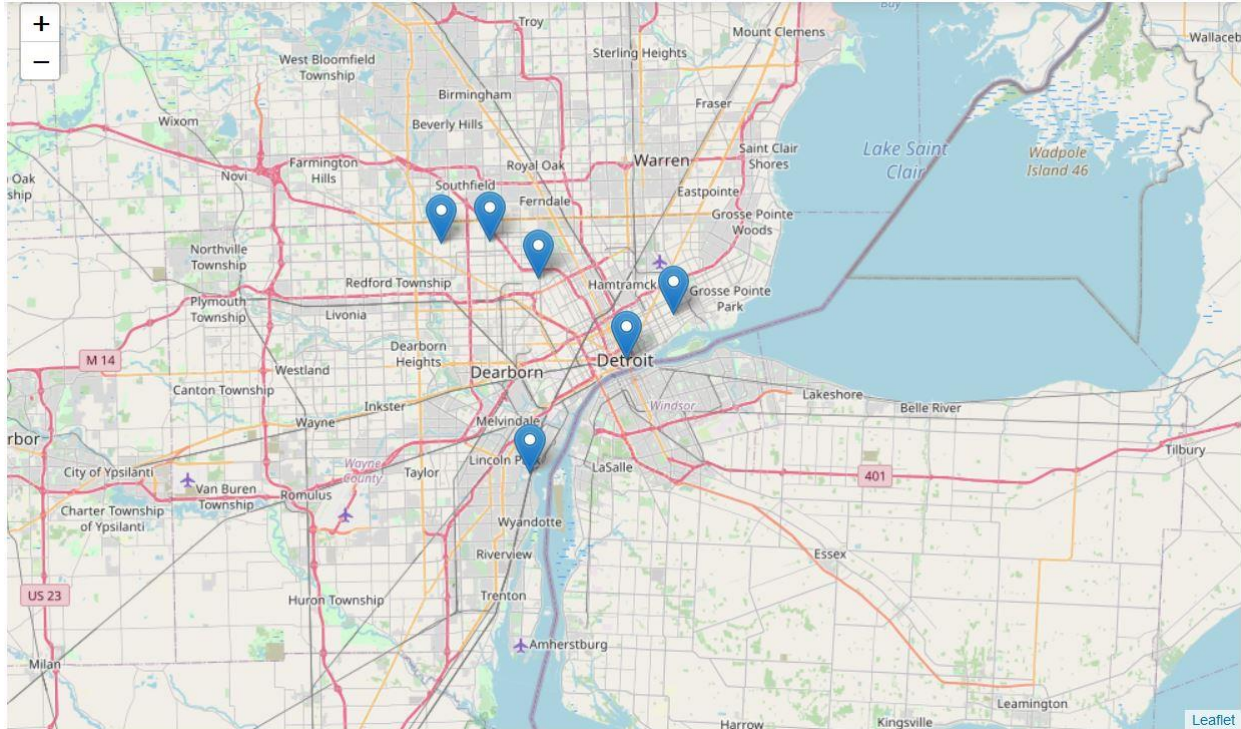
Snippet of Detroit Demographic Pandas DataFrame

2. Zip Code and Geocoordinate data of Detroit is obtained from public.opensoft.com API

Downloaded as a json file and process into a pandas DataFrame, cleaned and visualized using python geopy library and Folium.

	Zip Code	Latitude	Longitude
0	48238	42.396736	-83.141520
1	48297	42.239933	-83.150823
2	48269	42.239933	-83.150823
3	48258	42.239933	-83.150823
4	48272	42.239933	-83.150823
5	48214	42.368137	-82.993140
6	48219	42.425236	-83.249020
7	48267	42.239933	-83.150823
8	48235	42.427636	-83.195470
9	48266	42.239933	-83.150823
11	48201	42.331429	-83.045753

Zip code, latitude and longitude frame

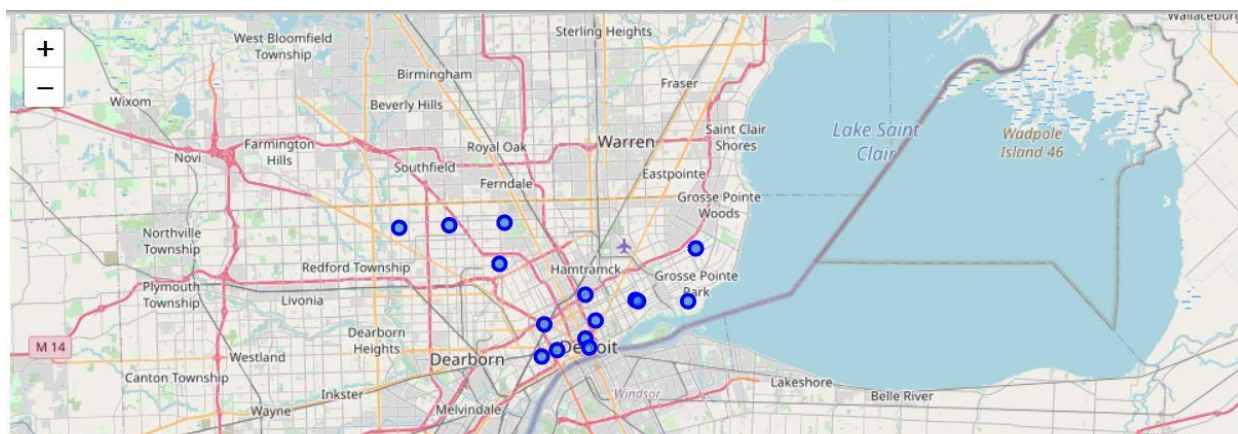


Folium map of Detroit with coordinates and zip codes of the DataFrame

- Additional data of Detroit neighborhoods was collected from a Google Maps search and loaded into a pandas DataFrame. Note: A csv file was created independently since no file of the desired information is available for download.

	Zip Code	Neighborhood	Latitude	Longitude
0	48209	Mexicantown	42.324493	-83.096043
1	48201	Corktown	42.329569	-83.079564
2	48208	Core City	42.349617	-83.093983
3	48226	Downtown Detroit	42.338199	-83.049694
4	48207	Eastern Market	42.352155	-83.039051
5	48211	Poletown	42.372449	-83.049351
6	48214	Indian Village	42.368391	-82.995793
7	48215	Jefferson Chalmers	42.368137	-82.939488
8	48224	East English Village	42.408963	-82.931591
9	48221	University District	42.429746	-83.136212

4. The Zip Code DataFrame and Neighborhood DataFrame were merged, cleaned and then visualized using Folium.



Exploring Neighborhoods and Venues with Foursquare

A list of venues by neighborhood was obtained using the Foursquare API and processed into a pandas DataFrame. An examination of and count of the venues by neighborhood was then performed.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Mexicantown	42.324493	-83.096043	El Club	42.321434	-83.093511	Music Venue
1	Mexicantown	42.324493	-83.096043	Mexicantown Bakery	42.320815	-83.094930	Bakery
2	Mexicantown	42.324493	-83.096043	Taqueria El Rey	42.319166	-83.098925	Mexican Restaurant
3	Mexicantown	42.324493	-83.096043	Flowers Of Vietnam	42.320109	-83.096501	Vietnamese Restaurant
4	Mexicantown	42.324493	-83.096043	La Jalisciense Taqueria	42.321991	-83.091155	Taco Place

DataFrame of Venues obtained from Foursquare

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
City	141	141	141	141	141	141
Core City	19	19	19	19	19	19
Corktown	52	52	52	52	52	52
Downtown Detroit	100	100	100	100	100	100
East English Village	32	32	32	32	32	32
Eastern Market	64	64	64	64	64	64
Indian Village	6	6	6	6	6	6
Jefferson Chalmers	19	19	19	19	19	19
Mexicantown	35	35	35	35	35	35
Poletown	11	11	11	11	11	11

DataFrame of Venues Counts

Analyzing the Neighborhoods and Venues

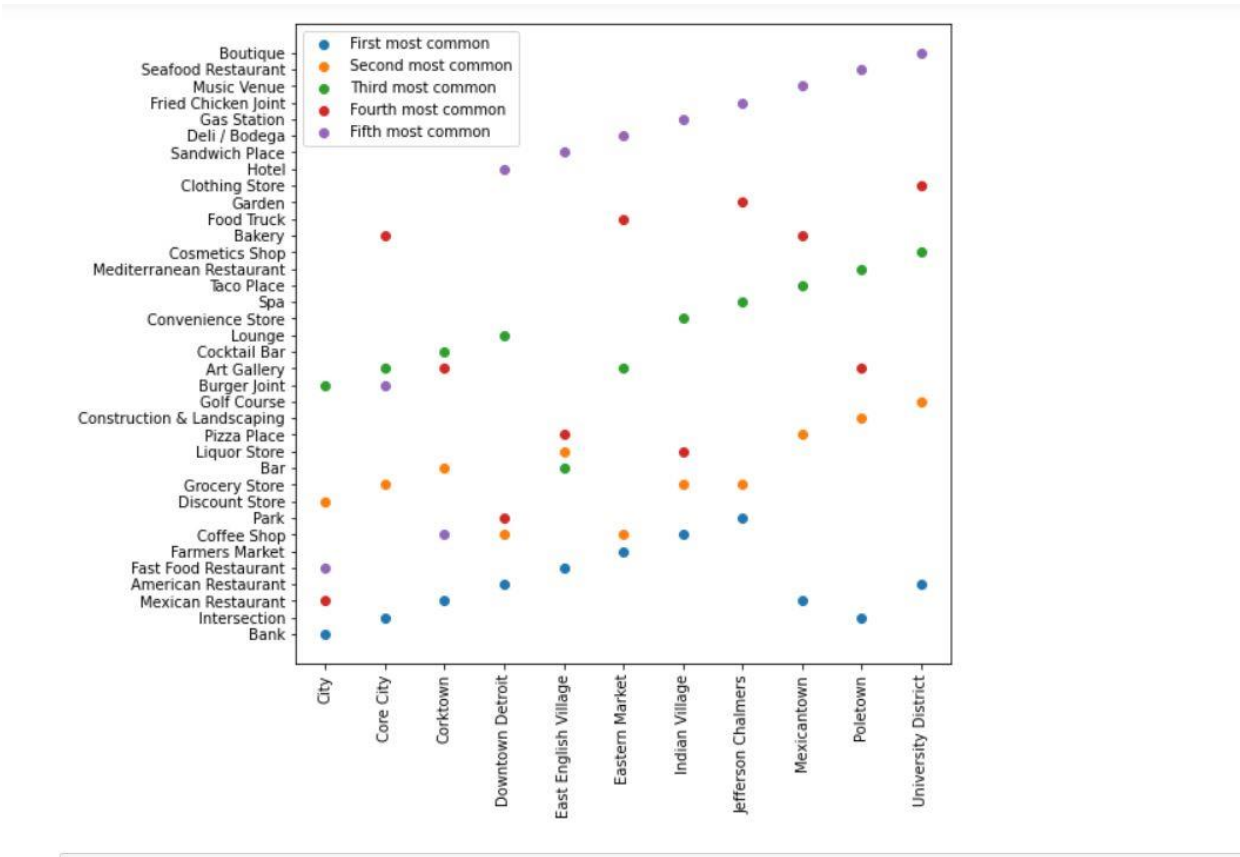
Categorical variables were first transformed using one-hot-encoding and the mean of the frequency of the occurrence of the categories was calculated.

	Neighborhood	Accessories Store	American Restaurant	Antique Shop	Art Gallery	Asian Restaurant	Athletics & Sports	Auto Workshop	Automotive Shop	BBQ Joint	Bakery	Bank	Bar	Baseball Stadium	Bre
0	City	0.0	0.014184	0.000000	0.014184	0.007092	0.000000	0.000000	0.000000	0.000000	0.000000	0.06383	0.021277	0.007092	0.0
1	Core City	0.0	0.000000	0.052632	0.105263	0.000000	0.000000	0.052632	0.000000	0.000000	0.105263	0.000000	0.000000	0.000000	0.0
2	Corktown	0.0	0.000000	0.019231	0.038462	0.019231	0.019231	0.000000	0.019231	0.019231	0.038462	0.000000	0.076923	0.000000	0.0
3	Downtown Detroit	0.0	0.060000	0.000000	0.010000	0.000000	0.000000	0.000000	0.000000	0.010000	0.000000	0.010000	0.040000	0.010000	0.0
4	East English Village	0.0	0.031250	0.000000	0.000000	0.000000	0.031250	0.000000	0.000000	0.031250	0.000000	0.000000	0.093750	0.000000	0.0

The venues were then sorted in descending order and a new DataFrame was created showing the top ten most common venues for each neighborhood. A Scatterplot was used for visualization.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	City	Bank	Discount Store	Burger Joint	Mexican Restaurant	Fast Food Restaurant	Pharmacy	Park	Coffee Shop	Hotel	Steakhouse
1	Core City	Intersection	Grocery Store	Art Gallery	Bakery	Burger Joint	Skate Park	Bus Station	Mediterranean Restaurant	Check Cashing Service	Men's Store
2	Corktown	Mexican Restaurant	Bar	Cocktail Bar	Art Gallery	Coffee Shop	Bakery	New American Restaurant	Pub	Historic Site	Sandwich Place
3	Downtown Detroit	American Restaurant	Coffee Shop	Lounge	Park	Hotel	Steakhouse	Bar	New American Restaurant	Dessert Shop	Burger Joint
4	East English Village	Fast Food Restaurant	Liquor Store	Bar	Pizza Place	Sandwich Place	Fried Chicken Joint	Home Service	Fishing Spot	Deli / Bodega	Diner

First five rows of the Top 10 Venues DataFrame

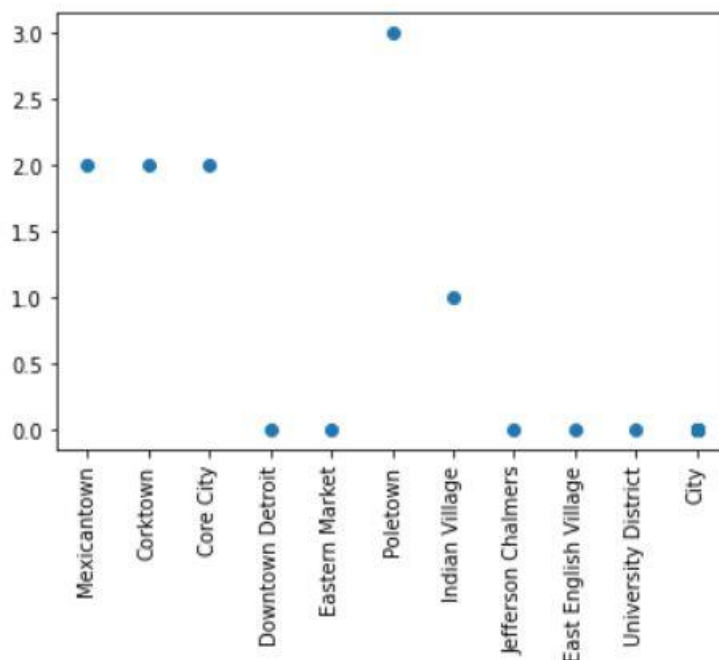


Scatterplot of the 5 most common venues

Performing K-Means clustering for segmentation

The K-Means algorithm was run individually through two to nine clusters, then visualized using scatterplots to estimate the best k. Calculating and reviewing the sum of squared error and plotting with the elbow method was used as well. It was determined that using four clusters was best and a Folium map was used for visualization.

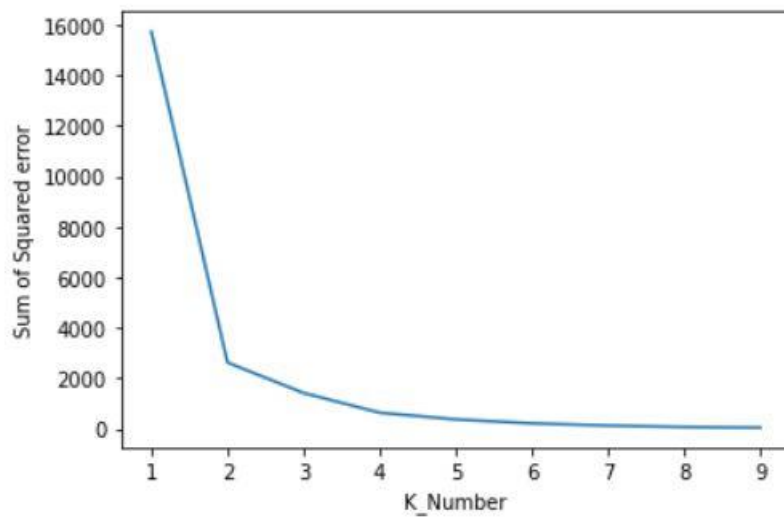
```
([0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10],  
 <a list of 11 Text major ticklabel objects>)
```



Plot of four clusters

```
[15733.238095238099,  
2625.2333333333336,  
1418.1111111111113,  
637.8111111111111,  
371.20000000000005,  
216.4142857142857,  
128.21428571428572,  
71.33333333333333,  
46.33333333333333]
```

A look at the Sum of Squared Error



Elbow Plot

Examining the Clusters

First Cluster

Cluster label 0

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	Downtown Detroit	American Restaurant	Coffee Shop	Lounge	Park	Hotel	Steakhouse	Bar	New American Restaurant	Dessert Shop	Burger Joint
4	Eastern Market	Farmers Market	Coffee Shop	Art Gallery	Food Truck	Deli / Bodega	Seafood Restaurant	Steakhouse	Food	Clothing Store	Antique Shop
7	Jefferson Chalmers	Park	Grocery Store	Spa	Garden	Fried Chicken Joint	Liquor Store	Discount Store	Coffee Shop	Chinese Restaurant	Carpet Store
8	East English Village	Fast Food Restaurant	Liquor Store	Bar	Pizza Place	Sandwich Place	Fried Chicken Joint	Home Service	Fishing Spot	Deli / Bodega	Diner
9	University District	American Restaurant	Golf Course	Cosmetics Shop	Clothing Store	Boutique	Insurance Office	Intersection	Mediterranean Restaurant	Business Service	Shoe Store
10	City	Bank	Discount Store	Burger Joint	Mexican Restaurant	Fast Food Restaurant	Pharmacy	Park	Coffee Shop	Hotel	Steakhouse
			Discount		Mexican	Fast Food					

Second cluster

Cluster label 1

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
6	Indian Village	Coffee Shop	Grocery Store	Convenience Store	Liquor Store	Gas Station	Food	Greek Restaurant	Financial or Legal Service	Construction & Landscaping	Harbor / Marina

Third Cluster

Cluster label 2

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Mexicantown	Mexican Restaurant	Pizza Place	Taco Place	Bakery	Music Venue	Restaurant	Sandwich Place	Diner	Dessert Shop	Deli / Bodega
1	Corktown	Mexican Restaurant	Bar	Cocktail Bar	Art Gallery	Coffee Shop	Bakery	New American Restaurant	Pub	Historic Site	Sandwich Place
2	Core City	Intersection	Grocery Store	Art Gallery	Bakery	Burger Joint	Skate Park	Bus Station	Mediterranean Restaurant	Check Cashing Service	Men's Store

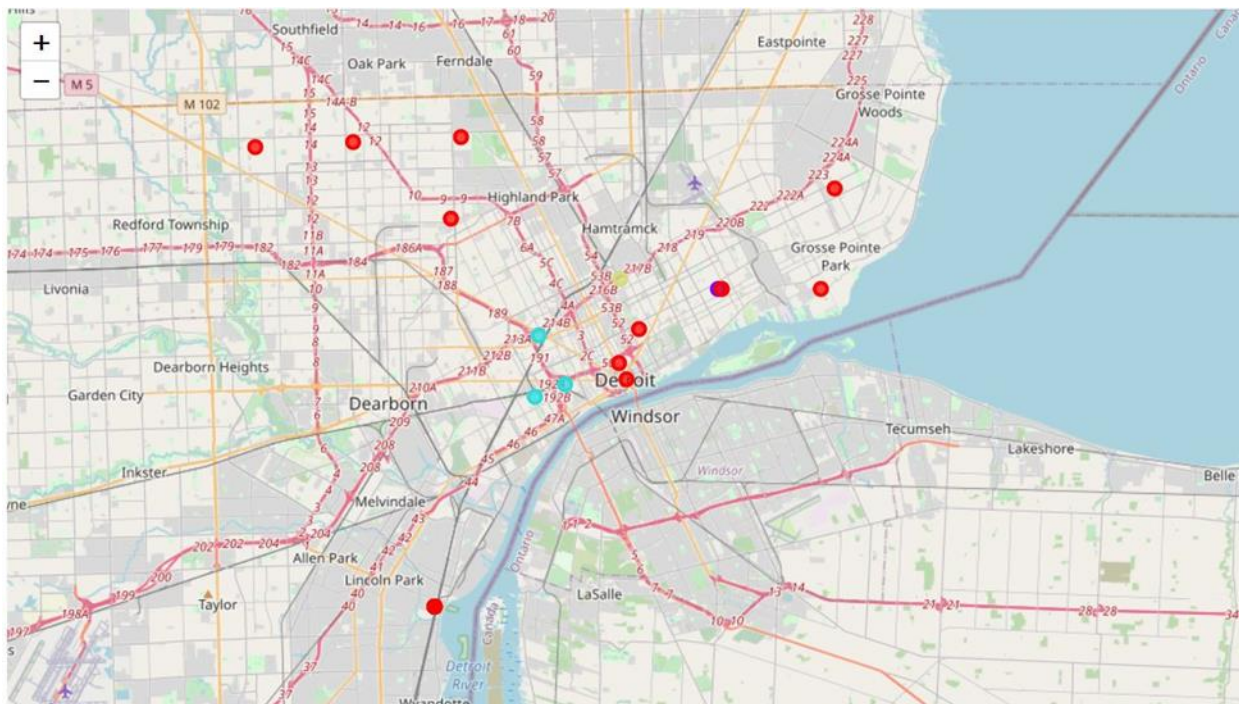
Fourth Cluster

Cluster label 3

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
5	Poletown	Intersection	Construction & Landscaping	Mediterranean Restaurant	Art Gallery	Seafood Restaurant	Speakeasy	Food	Print Shop	Market	Restaurant

Results

The results of applying the K-Means algorithm to the selected neighborhoods in Detroit show the distribution of venues in four clusters.



1. Cluster 0 in red, with the largest number of dining establishments, highlights neighborhoods mainly on the periphery of Detroit with the exception of the Downtown, a designated City area, and the Eastern Market Neighborhood.

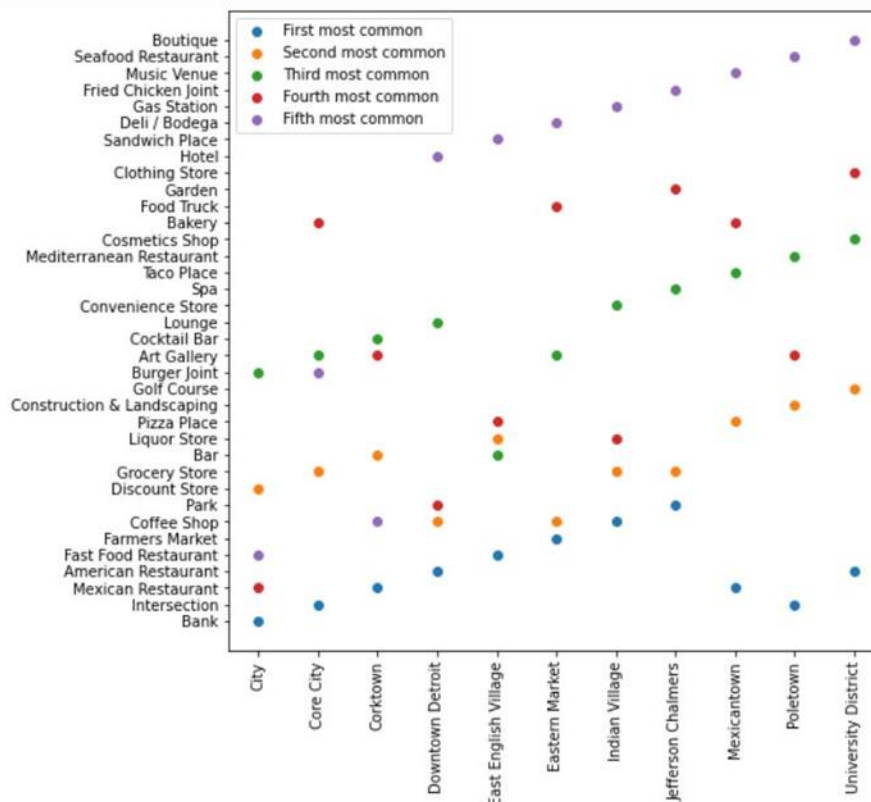
2. Cluster 1 in purple, highlights the Indian Village neighborhood with sparse dining opportunities.
3. Cluster 2 in light blue, highlights the Mexicantown, Corktown and Core City neighborhoods. These neighborhoods have a good number of dining establishments.
4. Cluster 3 in light brown, highlights the Poletown neighborhood, having among the fewest dining venues.

Discussion

For the start-up restaurateur or the seasoned investor, a first understanding of the big picture is a must. What are the popular dining establishments in each neighborhood? What are the most promising neighborhoods for development or investment regarding unmet demand and competition? This analysis addresses these questions and begins to point a way for stakeholder decisions toward maximizing success.

So, what are the most popular dining establishments in each of our selected neighborhoods? To help answer that question, let's have another look at the scatterplot used in the beginning of our analysis.

Scatterplot of the top five Neighborhood Venues



Listing the top five dining establishments for each neighborhood we see:

City – Mexican restaurant, Fast food/burger joint

Core City - Burger Joint

Corktown – Mexican restaurant, Coffee Shop

Downtown Detroit – American restaurant, Coffee shop

East English Village – Fast food, Deli

Indian Village – Coffee shop

Jefferson Chalmers – Fried Chicken Joint

Mexicantown – Mexican restaurant, mobile food truck

Poletown – Mediterranean restaurant, Seafood restaurant

University District – American restaurant, Burger joint

We can see from the scatterplot that the most popular type of dining establishment among the Detroit neighborhoods is the Fast food restaurant or Burger joint. Next in popularity are Mexican restaurants and coffee shops, then American style restaurants and Delis.

For an indication of what might be the most promising neighborhoods for development or investment with reasonable competition we can review the neighborhood clusters.

In Cluster 0, the most widespread and populous cluster, encompassing mostly the periphery of Detroit, the Jefferson Chalmers neighborhood appears to have some potential for development owing to few dining establishments and none of the most popular in its top five.

Cluster 1, Indian Village and Cluster 3, Poletown, also appear to be good candidates for further development due to having few dining establishments and to their locations.

Conclusion

Using a combination of datasets obtained from or created independently from the US Census Bureau, Foursquare Developer API, public.opendatasoft.com and Google Maps search results, selected neighborhoods in Detroit were analyzed for their potential as areas for opening or investing in a restaurant. Through a process of statistical description, visualizations and K-Means clustering the most popular types of dining establishments in each neighborhood were determined, and the most promising neighborhoods for development were suggested.

Stakeholder business knowledge combined with a review of the Detroit demographic data and the neighborhood analysis, points a way toward decisions for maximizing success.

